SUPPORTING INFORMATION

Evolutionary dynamics of sex-biased genes expressed in cricket brains and gonads

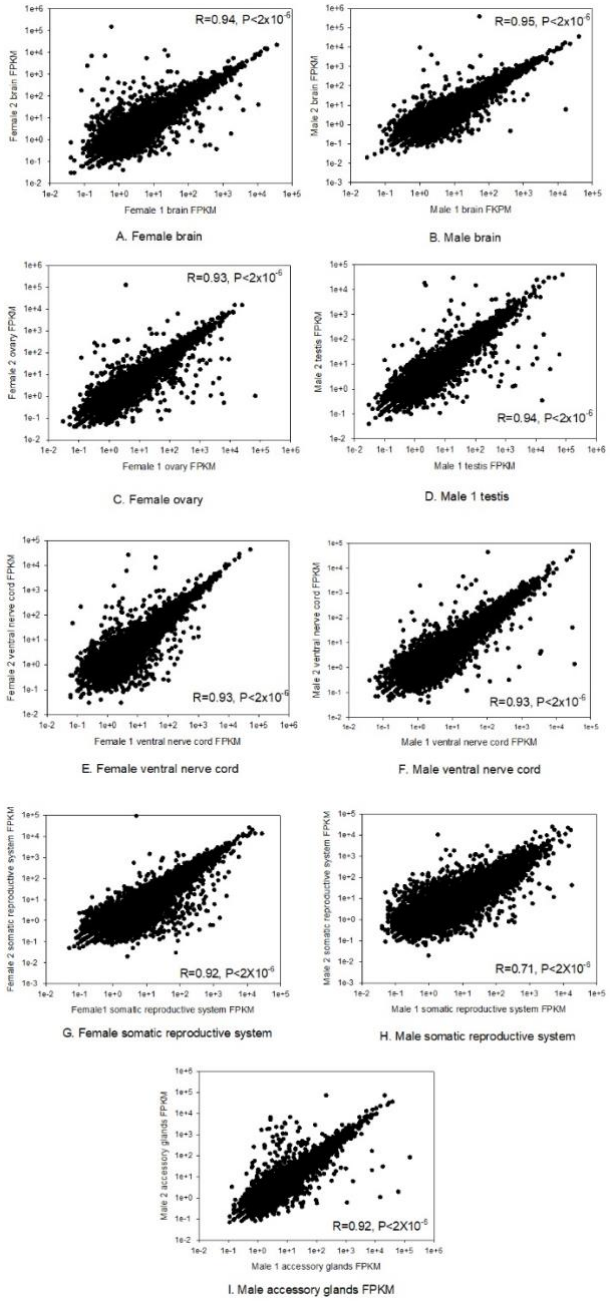

Fig. S1. The Spearman correlation (R) in FPKM across all all 15,539 genes in *G. bimaculatus* for each of the tissues under study. A) female brain; B) male brain; C) ovary; D) testis; E) female ventral nerve cord; F) male ventral nerve cord; G) female somatic reproductive system; H) male somatic reproductive system; I) male accessory glands.

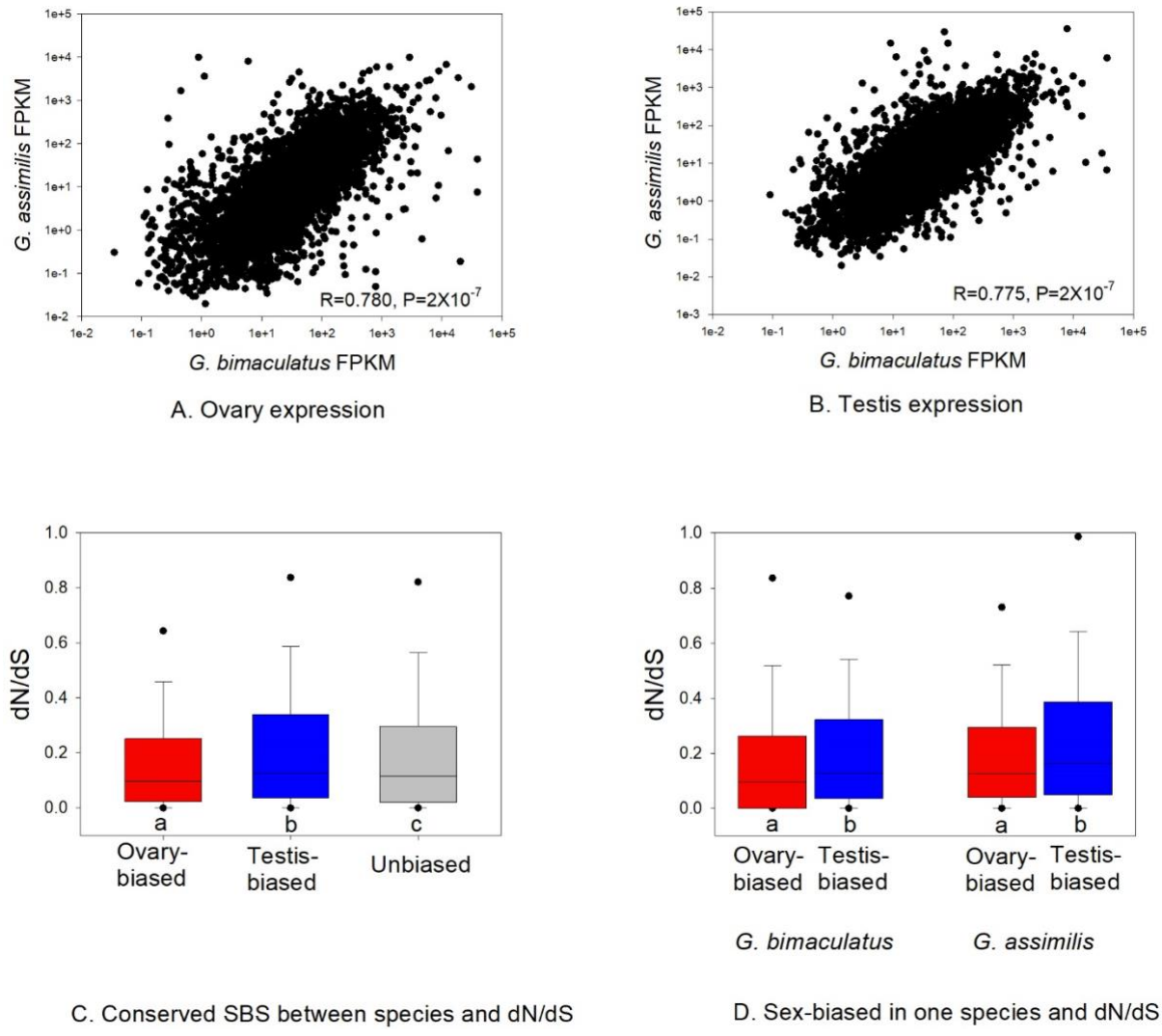

Fig. S2. Expression of genes in *G. bimaculatus* and *G. assimilis* for A) ovaries and B) testes (Spearman's R and P are shown; expression per gene is the average across samples per species). C) Box plots of dN/dS for genes that have the same sex-biased status (SBS) in both *G. bimaculatus* and *G. assimilis* and; D) Box plots of dN/dS for genes that are ovary-biased or testis-biased in only one species. Different letters under bars in C and each pair of bars in D indicate MWU-P<0.05.

Table S1. The RNA-seq datasets for each of the male and female tissue types under study for *G. bimaculatus*. The number of reads (single-end) before and after trimming with BBduk (https://jgi.doe.gov/data-and-tools/bbtools/) is shown. The data are available at the Short Read Archive (SRA) under the project identifier PRJNXXXXXX.

			No. Reads	
Sample	**Tissue**	**Sample Name**	**Before trimming**	**After trimming**
Male 1	Accessory gland	AK-28_S6.R1	8,519,999	8,455,381
	Brain	AK-25_S3.R1	10,927,264	10,543,501
	Somatic reproductive system	SHC-18_S14.R1	32,497,283	32,430,843
	Testes	SHC-17_S13.R1	19,928,912	19,751,731
	Ventral nerve cord	AK-26_S4.R1	11,488,521	11,140,299
Male 2	Accessory gland	AK-35_S13.R1	15,110,718	14,973,668
	Brain	AK-32_S10.R1	18,039,328	17,850,399
	Somatic reproductive system	AK-31_S9.R1	11,993,680	11,702,596
	Testes	AK-30_S8.R1	13,672,147	13,529,248
	Ventral nerve cord	AK-33_S11.R1	11,677,747	11,445,159
Female 1	Brain	AK-39_S17.R1	13,920,966	13,750,206
	Ovary	AK-37_S15.R1	21,725,208	21,128,416
	Somatic reproductive system	AK-38_S16.R1	13,870,827	13,718,497
	Ventral nerve cord	AK-40_S18.R1	12,599,661	12,341,413
Female 2	Brain	AK-45_S23.R1	19,312,301	19,036,974
	Ovary	AK-43_S21.R1	27,627,122	27,049,583
	Somatic reproductive system	AK-44_S22.R1	11,688,814	11,539,571
	Ventral nerve cord	AK-46_S24.R1	13,591,143	13,143,568

Table S2. The RNA-seq datasets for each of the male and female tissue types under study for *G. assimilis*. The number of reads (single-end) before and after trimming with BBduk is shown. All tissue samples were used for the RNA-seq assembly, and * indicates those samples used for gonadal expression (Exp.) analyses for comparison to *G. bimaculatus* tissue samples. The data are available at the Short Read Archive (SRA) under the project identifier PRJNAXXXXXX.

Sample	Tissue[a]	File Name	No. Reads		Exp.
			Before trimming	After trimming	
Male 1	Accessory gland	AK_1-6_S6_R1_001	7,232,269	6,927,036	
	Brain	AK_1-3_S3_R1_001	414,137	397,903	
	Carcass	AK_1-5_S5_R1_001	12,162,122	11,909,287	
	Somatic reproductive system	AK_1-2_S2_R1_001	9,786,892	9,218,317	
	Testes	AK_1-1_S1_R1_001	15,489,005	14,677,775	
	Ventral nerve cord	AK_1-4_S4_R1_001	8,085,561	7,778,520	
Male 2	Accessory gland	AK_1-12_S12_R1_001	9,684,938	9,395,491	
	Brain	AK_1-9_S9_R1_001	10,251,515	9,935,844	
	Carcass	AK_1-11_S11_R1_001	13,140,136	12,739,689	
	Somatic reproductive system	AK_1-8_S8_R1_001	9,038,153	8,700,106	
	Testes	AK_1-7_S7_R1_001	15,015,495	14,252,214	*
	Ventral nerve cord	AK_1-10_S10_R1_001	159,249,632	153,314,813[b]	
Male 3	Accessory gland	AK_1-18_S18_R1_001	9,301,332	8,771,926	
	Brain	AK_1-15_S15_R1_001	8,556,018	8,269,869	
	Carcass	AK_1-17_S17_R1_001	11,258,971	11,016,352	
	Somatic reproductive system	AK_1-14_S14_R1_001	11,468,584	11,055,961	
	Testes	AK_1-13_S13_R1_001	11,082,626	10,637,566	*
	Ventral nerve cord	AK_1-16_S16_R1_001	9,501,325	9,141,580	
Female 1	Brain	AK_1-21_S21_R1_001	12,314,902	11,529,951	
	Carcass	AK_1-23_S23_R1_001	10,318,471	9,965,655	
	Ovary	AK_1-19_S19_R1_001	12,968,675	12,330,995	*

	Somatic reproductive system	AK_1-20_S20_R1_001	20,180,007	19,613,713	
	Ventral nerve cord	AK_1-22_S22_R1_001	13,818,212	13,322,784	
Female 2	Brain	AK_1-26_S26_R1_001	10,596,275	10,191,182	
	Carcass	AK_1-28_S28_R1_001	9,471,179	8,987,504	
	Ovary	AK_1-24_S24_R1_001	14,894,350	14,584,072	*
	Somatic reproductive system	AK_1-25_S25_R1_001	10,705,738	10,183,713	
	Ventral nerve cord	AK_1-27_S27_R1_001	10,108,946	9,733,477	
Female 3	Brain	AK_1-31_S31_R1_001	9,543,257	9,388,801	
	Carcass	AK_1-33_S33_R1_001	14,562,167	14,279,995	
	Ovary	AK_1-29_S29_R1_001	10,900,114	10,725,546	
	Somatic reproductive system	AK_1-30_S30_R1_001	9,846,659	9,641,174	
	Ventral nerve cord	AK_1-32_S32_R1_001	7,954,359	7,795,480	

[a] The carcass consists of body and head (minus the legs, wings, antennae, gut, Malpighian tubules, gonad and somatic reproductive system, ventral nerve cord, and brain); [b]The sample was divided into tenths and one-tenth of the reads were randomly used for assembly and expression analysis to approximately match the read number of other samples.

Table S3. The number of genes in *G. bimaculatus* under study for dN/dS (those with between-species orthologs) that had female- or male-biased expression in one (of four) paired tissue types and were unbiased in expression for the three remaining tissue types (tissue-specific sex bias, TSSB). Genes with shared sex-biased status in two of four tissues (and unbiased in two) or other types of variation in SBS are also shown.

	Number of Genes	
Sex-biased one tissue and unbiased in three tissues (TSSB, N=3,375	Female-biased	Male-biased
Gonad	1,858	1,055
Somatic reproductive system	113	126
Brain	6	16
Ventral nerve cord	82	119
Shared status in two tissues N=226	Female-biased	Male-biased
Gonad and somatic reproductive system	37	67
Gonad and brain	3	7
Gonad and ventral nerve cord	56	44
Brain and somatic reproductive system	1	6
Brain and ventral nerve cord	1	4
Other status		Other status
Universally unbiased		3,449
Variable: multiple sex-biased statuses among tissues		170
Total number of genes in all categories		7,220

Table S4. The top GO functional classifications of genes with sex-biased brain expression (all genes regardless of interspecies orthologs, N values in Fig. 2). Annotation was determined in DAVID (Huang da et al., 2009) using *D. melanogaster* orthologs to *G. bimaculatus* genes. Functions are ranked by the percentage of genes matching its classification. P-values indicate the enrichment of genes involved in each function with lower values indicating greater enrichment.

Female-biased brain genes	**Percent of Genes**	**P-value**
Positive regulation of transcription from RNA polymerase II promoter	12.1	1.10E-02
Sensory perception of pain	12.1	3.40E-02
Transcription, DNA-templated	12.1	4.20E-02
Male-biased brain genes		
Proteolysis	11.3	2.90E-04
Calcium ion binding	8.5	6.70E-03
Neuron remodeling	4.2	2.10E-02

Table S5. The top GO functional classifications of genes with universally unbiased expression across all four paired male-female tissue types under study (those with interspecies orthologs for study). Annotation was determined in DAVID (Huang da et al., 2009) using *D. melanogaster* orthologs to *G. bimaculatus* genes. Functions are ranked by the percentage of genes matching its classification and those categories representing 5% or more of genes are shown. P-values indicate the enrichment of genes involved in each function with lower values indicating greater enrichment.

GO keyword	Percent of genes	P-value
Coiled coil	16.1	5.20E-16
Hydrolase	12.9	4.00E-03
Transferase	12.4	2.40E-21
Nucleus	11.2	8.30E-10
Phosphoprotein	11	3.40E-22
Metal-binding	10.8	6.00E-14
Nucleotide-binding	8.3	2.50E-15
Alternative splicing	8.1	4.40E-20
Cytoplasm	8	1.00E-17
ATP-binding	7	3.00E-16
Developmental protein	6.7	5.20E-11
Zinc	6.6	1.20E-09
Transport	5.1	1.70E-03
Oxidoreductase	5.0	4.30E-02

Table S6. The 30 genes that were expressed specifically in the male accessory glands and not in any of the eight other (male and female) tissue types in *G. bimaculatus*. The genes that had high confidence ortholog matches between *G. bimaculatus* and *G. assimilis* are shown (N=1, indicated by *), as well as those with putative orthologs identified in *D. melanogaster* (N=7), which had less strict criteria (for identification of a match for gene ontology purposes; See Methods). Gene functions were predicted using DAVID and *D. melanogaster* gene identifiers (Huang da et al., 2009).

Row	*G. bimaculatus* gene	*D. melanogaster* match	Gene	GO putative functions	Expression (FPKM)	dN/dS
1	GBI_11980-RA*	FBgn0035154	CG3344	Peptidase S10, serine carboxypeptidase	19.46	0.1039
2	GBI_06110-RA	FBgn0035781	CG8560	proteolysis	9.60	-
3	GBI_00239-RA	FBgn0039084	CG10175	neuron cell-cell adhesion, synaptic transmission	0.83	-
4	GBI_14669-RA	FBgn0259215	*Ionotropic receptor 93a (Ir93a)*	detection of chemical stimulus sensory perception	0.70	-
5	GBI_18443-RA	FBgn0035476	CG12766	oxidation-reduction process	0.34	-
6	GBI_06560-RA	FBgn0024288	*Sox100B*	male gonad development	0.27	-
7	GBI_04401-RA	FBgn0060296	*painless(pain)*	copulation, male courtship behavior, olfactory	0.08	-
8	GBI_08022-RA	No match	-	-	347.25	-
9	GBI_00292-RA	No match	-	-	120.78	-
10	GBI_13609-RA	No match	-	-	41.80	-
11	GBI_13241-RA	No match	-	-	20.21	-
12	GBI_14160-RA	No match	-	-	16.75	-
13	GBI_01608-RA	No match	-	-	15.96	-
14	GBI_05352-RA	No match	-	-	8.13	-
15	GBI_06938-RA	No match	-	-	5.72	-
16	GBI_21228-RA	No match	-	-	5.29	-
17	GBI_09340-RA	No match	-	-	3.41	-
18	GBI_17179-RA	No match	-	-	3.39	-
19	GBI_18175-RA	No match	-	-	1.42	-
20	GBI_03401-RA	No match	-	-	1.21	-
21	GBI_06890-RA	No match	-	-	1.15	-
22	GBI_06913-RA	No match	-	-	1.08	-

23	GBI_02473-RA	No match	-	-	0.92	-
24	GBI_02976-RA	No match	-	-	0.74	-
25	GBI_05550-RA	No match	-	-	0.6	-
26	GBI_00099-RA	No match	-	-	0.49	-
27	GBI_11938-RA	No match	-	-	0.35	-
28	GBI_16483-RA	No match	-	-	0.35	-
29	GBI_07340-RA	No match	-	-	0.29	-
30	GBI_04232-RA	No match	-	-	>0[a]	-

[a] Highly variable FPKM among individuals and thus denoted as >0 FPKM.

Table S7. The 134 seminal fluid protein (SFPs) genes for the species *D. melanogaster* from Sepil et al. 2019 (Sepil et al., 2019) and their best BLASTX matches in the 15,539 genes of *G. bimaculatus*. Due to the extended phylogenetic distance between species, the list shows all putative orthologs identified using single forward BLASTX of *G. bimaculatus* to *D. melanogaster* using BLASTX (e<0.001). Results for those contained within the 7,220 genes with high confidence orthologs between *G. bimaculatus* and *G. assimilis* are shown in Table 4 within the main text.

Row	***D. melanogaster* ID**	**SFP gene name**	**Match in *G. bimaculatus***
1	FBgn0002855	*Acp26Aa*	-
2	FBgn0002856	*Acp26Ab*	-
3	FBgn0015583	*Acp29AB*	-
4	FBgn0267327	*Acp33A*	-
5	FBgn0011559	*Acp36DE*	-
6	FBgn0034152	*Acp53C14a*	-
7	FBgn0034153	*Acp53C14b*	-
8	FBgn0053530	*Acp53C14c*	-
9	FBgn0015584	*Acp53Ea*	-
10	FBgn0020509	*Acp62F*	-
11	FBgn0015585	*Acp63F*	-
12	FBgn0015586	*Acp76A*	-
13	FBgn0003884	*alphaTub84B*	GBI_00369-RA
14	FBgn0050488	*antr*	-
15	FBgn0039598	*aqrs*	-
16	FBgn0003889	*betaTub85D*	-
17	FBgn0047334	*BG642312*	-
18	FBgn0054002	*BP1025*	-
19	FBgn0038014	CG10041	GBI_00322-RA*
20	FBgn0260766	CG42564	-
21	FBgn0038395	CG10407	GBI_00641-RA*
22	FBgn0037039	CG10587	-
23	FBgn0032853	CG10651	-
24	FBgn0032843	CG10730	-
25	FBgn0037038	CG11037	-
26	FBgn0033164	CG11112	-
27	FBgn0038067	CG11598	-
28	FBgn0038069	CG11608	-
29	FBgn0250847	CG14034	-
30	FBgn0034417	CG15117	GBI_01865-RA
31	FBgn0031617	CG15635	-
32	FBgn0030643	CG15641	-

33	FBgn0033167	CG1701	-
34	FBgn0051872	CG31872	-
35	FBgn0265264	CG17097	-
36	FBgn0250841	CG17242	-
37	FBgn0032868	CG17472	-
38	FBgn0250842	CG17575	-
39	FBgn0038919	*Qsox2*	-
40	FBgn0037433	CG17919	GBI_09042-RB
41	FBgn0034512	CG18067	-
42	FBgn0036837	CG18135	-
43	FBgn0043825	CG18284	-
44	FBgn0034753	CG2852	GBI_11684-RA
45	FBgn0050395	CG30395	-
46	FBgn0050486	CG30486	-
47	FBgn0051418	CG31418	-
48	FBgn0051419	CG31419	-
49	FBgn0051515	CG31515	-
50	FBgn0051659	CG31659	-
51	FBgn0051680	CG31680	-
52	FBgn0051704	CG31704	-
53	FBgn0032122	CG31883	-
54	FBgn0052833	CG32833	-
55	FBgn0054002	CG34002	-
56	FBgn0054033	CG34033	-
57	FBgn0054034	CG34034	-
58	FBgn0054051	CG34051	-
59	FBgn0083965	CG34129	-
60	FBgn0083966	CG34130	-
61	FBgn0260766	CG42564	-
62	FBgn0263024	CG43319	-
63	FBgn0034229	CG4847	-
64	FBgn0030828	CG5162	-
65	FBgn0036186	CG6071	-
66	FBgn0038918	*Qsox3*	-
67	FBgn0031746	CG9029	-
68	FBgn0035216	CG9168	-
69	FBgn0039597	CG9997	-
70	FBgn0004629	*Cys*	-
71	FBgn0250832	*Dup99B*	-
72	FBgn0004181	*Ebp*	-
73	FBgn0011694	*EbpII*	-
74	FBgn0000592	*Est-6*	GBI_00242-RA
75	FBgn0030932	*Ggt-1*	GBI_03406-RA

76	FBgn0041629	*Hexo2*	GBI_01177-RA
77	FBgn0040098	*lectin-29Ca*	-
78	FBgn0040097	*lectin-30A*	-
79	FBgn0040093	*lectin-46Ca*	-
80	FBgn0040092	*lectin-46Cb*	-
81	FBgn0028416	*Met75Ca*	-
82	FBgn0260745	*mfas*	GBI_04258-RA
83	FBgn0011668	*Mst57Da*	-
84	FBgn0011670	*Mst57Dc*	-
85	FBgn0053126	*NLaz*	GBI_14572-RA
86	FBgn0038198	*Npc2b*	GBI_06029-RA
87	FBgn0052190	*NUCB1*	GBI_02944-RA
88	FBgn0043539	*Obp22a*	-
89	FBgn0043530	*Obp51a*	-
90	FBgn0034471	*Obp56e*	GBI_19371-RA
91	FBgn0043533	*Obp56f*	-
92	FBgn0034474	*Obp56g*	GBI_14450-RA
93	FBgn0043532	*Obp56i*	-
94	FBgn0283509	*Phm*	GBI_06121-RA
95	FBgn0069354	*Porin2*	-
96	FBgn0030362	*regucalcin*	GBI_08029-RA
97	FBgn0033868	*S-Lap7*	-
98	FBgn0028944	*Semp1*	-
99	FBgn0037036	*Sems*	-
100	FBgn0259949	*Sfp23F*	-
101	FBgn0259951	*Sfp24Ba*	-
102	FBgn0259952	*Sfp24Bb*	-
103	FBgn0261054	*Sfp24Bc*	-
104	FBgn0259953	*Sfp24Bd*	-
105	FBgn0259956	*Sfp24C1*	-
106	FBgn0259958	*Sfp24F*	-
107	FBgn0259959	*Sfp26Ac*	-
108	FBgn0261055	*Sfp26Ad*	-
109	FBgn0259964	*Sfp33A3*	-
110	FBgn0259965	*Sfp35C*	-
111	FBgn0261058	*Sfp38D*	-
112	FBgn0259966	*Sfp51E*	-
113	FBgn0259969	*Sfp65A*	-
114	FBgn0259970	*Sfp70A4*	-
115	FBgn0261059	*Sfp78E*	-
116	FBgn0259975	*Sfp87B*	-
117	FBgn0003034	*SP*	-
118	FBgn0037038	*SP191*	-

119	FBgn0083141	*Spn28B*	-
120	FBgn0028987	*Spn28F*	GBI_00301-RB
121	FBgn0028986	*Spn38F*	GBI_05353-RD
122	FBgn0028988	*Spn42Dd*	-
123	FBgn0052203	*Spn75F*	-
124	FBgn0036969	*Spn77Bb*	-
125	FBgn0036970	*Spn77Bc*	-
126	FBgn0051413	*Qsox4*	-
127	FBgn0030589	CG9519	-
128	FBgn0085476	CG34447	-
129	FBgn0029804	CG3097	GBI_21205-RA*
130	FBgn0262621	CG43145	-
131	FBgn0053121	*Spn28Db*	-
132	FBgn0035042	CG3640	-
133	FBgn0083938	*BG642163*	-
134	FBgn0262571	CG43111	-

* In addition to the single-direction BLASTX, a reciprocal BLASTX between *G. bimaculatus* and *D. melanogaster* was conducted for SFP genes. Each match that did not have a best hit reciprocal match (not yielding the exact same match in the top three hits) is indicated by an asterisk (*)). Thus, these are lower confidence putative SFP orthologs in *G. bimaculatus*.

Table S8. The number of orthologs identified between *G. bimaculatus* (GB) and *G. assimilis* (GA) and the outgroup *L. kohalensis* (LK). Among the orthologs studied for GB-GA paired analysis, the number of genes also having an LK ortholog after three-way reciprocal BLASTX and after excluding genes with dN or dS>3 are shown (designated as high confidence). Branch-site analysis results including 2XΔL are shown for all genes studied and for sex-biased TSSB genes (or all genes for the brain) and universally unbiased genes. Genes not belonging to any of these categories were excluded (Table S3).

Identification of three-species orthologs	**All genes**	**Gonad-biased$_{TSSB}$**		**Somatic reproductive system-biased$_{TSSB}$**		**Ventral-biased$_{TSSB}$**		**Brain-biased$_{ALL}$**		**Universally unbiased**
		Ovary-biased	Testis-biased	Female-biased	Male-biased	Female-biased	Male-biased	Female-biased	Male-biased	
N GB-GA-LK putative orthologs (BLASTX)	4,523[a]	1,300	597	62	59	49	53	15	15	2,171
N with dN and dS < 3 (High confidence)	1,933[a]	553	250	33	20	14	31	6	4	927
N branch-site test (2XΔlnL) P<0.05	220[a]	65	24	6	1	1	7	1	2	101
Percent of studied genes statistically significant	11.38%	11.75%	9.60%	18.18%	5.00%	7.14%	22.58%	NA[b]	NA[b]	10.90%

[a]The total is for all genes with orthologs, including some not belonging to any of the sub-categories. [b] Too few genes were available to study for a reliable estimate.

Text File S1

Assembly of *G. assimilis* RNA-seq data

To assess dN/dS, we compared the annotated genes in *G. bimaculatus* to the CDS list generated for *G. assimilis*. Applying Trinity and PlantTribes (see Methods) to the trimmed reads in Table S2, we obtained 33,089 non-redundant transcripts with a median and mean length of 540 bp and 784.3 bp respectively (standard error=30.3). The BUSCO score (Seppey et al., 2019) to the Arthropoda conserved gene set of 1,066 genes, showed 86.7% CDS had complete sequence matches, 8.6% were fragmented matches, and 4.7% were missing. The latter may represent gene losses in this species, and/or genes excluded from the assembly. Thus, this suggests high efficiency of the assembly spanning a major portion of arthropod genes. From this list, we used ORF predictor with *G. bimaculatus* CDS as a reference and BLASTX to identify *G. assimilis* CDS. We found 25,128 CDS (including isoforms) with a start codon and no unknown or ambiguous nucleotides, which were used for analyses. Reciprocal BLASTX of the 15,539 *G. bimaculatus* CDS to the *G. assimilis* CDS yielded 7,919 putative orthologs between the two species ($e<10^{-6}$ in both forward and reverse matches). Retaining only those putative ortholog matches that after alignment had both dN and dS values <1.5, and thus were unsaturated, yielded a total of 7,220 high confidence between-species orthologs that were used for all our dN/dS analyses.

Comparison of sex-biased gonadal expression in *G. bimaculatus* and *G. assimilis*

As described in our main text, our core target for expression analysis was *G. bimaculatus*, which has an assembled genome with complete or near complete CDS (Ylla et al., 2021), and *G. assimilis* was used primarily as a reference for assessment of protein divergence. Nonetheless, we assessed the degree of conservation of gene expression between species for the 7,220 genes with orthologs for the gonads (largest N values of all tissues, Table S3) between these two species. The results showed that gene expression in *G. assimilis* gonads was strongly correlated to that in *G. bimaculatu*s, with Spearman's R=0.780 and 0.775 ($P<2X10^{-7}$) for ovary and testis expression respectively (Fig. S2AB). In addition, 65.9 and 65.8% of all gonadally expressed genes (among the 7,220 with orthologs) that were defined as female- and male-biased in *G. bimaculatus* (ALL genes sex-biased in testis regardless of status in other tissues, N=2,043 and 1,225 respectively) had the same status in *G. assimilis*. This suggests substantial turnover in sex-biased status, a pattern observed for gonadal tissues in studied species of *Drosophila* (Assis et al., 2012, Whittle & Extavour, 2019b, Zhang et al., 2007, Harrison et al., 2015). Importantly, for genes with the same (conserved) sex-biased status in the two species, dN/dS was highest in testis-biased genes (median=0.127) and lower in unbiased (0.114), and ovary-biased (0.097) genes (MWU-tests P<0.05 for all paired contrasts) (Fig. S2C). Moreover, genes that were testis-biased in only one species (either *G. bimaculatus* or *G. assimilis*) and unbiased in the other species had elevated dN/dS values as compared to their ovary-biased counterparts (MWU-test P<0.05 for each contrast, Fig. S2D). Thus, the accelerated evolution of testis-biased genes is

robust to whether the sex-biased status is observed in one species, or both species, in this taxon. All our remaining analysis is using sex-biased genes from our annotated model *G. bimaculatus*.

Assessment of expression in male accessory glands and seminal fluid proteins

We considered the evolution of genes specifically linked to the male accessory glands in *G. bimaculatus*, including those defined as putative orthologs to *D. melanogaster* seminal fluid proteins (SFPs; see below paragraph (Sepil et al., 2019)). First, we took a broad approach to study all male accessory gland-specific genes identified using our RNA-seq dataset (Table S1). Prior study of two species of crickets (*G. firmus* and *G. pennsylvanicus*) identified transcripts from the male accessory glands or SFPs, whereby some were suggested to evolve rapidly (Andres et al., 2006, Andres et al., 2013). Herein, we have the advantages of large-scale RNA-seq data from multiple tissue types, and an annotated *G. bimaculatus* genome (N=15,539 CDS) (Ylla et al., 2021), to identify male-accessory gland-specific genes in this species. We report a total of 30 genes expressed in the male accessory glands with no expression (0 FPKM) in all eight other studied male and female tissues (Table S1).

Functional predictions of the 30 male accessory gland-specific genes using *D. melanogaster* orthologs (Table S6, e<0.001, see Methods) revealed seven genes with a match. Two of these *G. bimaculatus* genes are predicted orthologs of *painless* and *Sox100B*, which have functions in male reproduction in *D. melanogaster*; the former is involved in courtship and olfactory signalling (Table S6). Both genes were expressed at low levels (FPKM<1) in male accessory glands in *G. bimaculatus*. Only one of the 30 accessory gland specific genes had a match in the two *Gryllus* species (3.33.%, Table S6, which had very strict match criteria, see Methods). Several of the *G. bimaculatus* accessory gland genes with no *G assimilis* or *D. melanogaster* ortholog matches had relatively high expression levels (e.g., 16 to 347 FPKM; Table S6), and we speculate they could comprise orphan genes that have evolved essential male sexual functions specifically in *G. bimaculatus* (Tautz & Domazet-Loso, 2011, Whittle & Extavour, 2019a). Overall, the nearly complete lack of high confidence orthologs between *G. bimaculatus* and *G. assimilis* suggests there has been rapid evolution of male accessory gland specific genes resulting in similarity too low for ortholog detection using these methods. Alternatively, these results may reflect a history of some lineage-specific gene losses or gains of these rapidly changing genes (Tautz & Domazet-Loso, 2011, Haerty et al., 2007).

Seminal fluid proteins

Seminal fluid proteins (SFPs) play significant roles in sperm vitality, sperm storage in the female spermatheca after mating, and in fertilization (Sepil et al., 2019). In studied systems to date, which have preferentially focused on primates and *Drosophila*, genes described as SFPs have been found to evolve rapidly and/or adaptively (Haerty et al., 2007, Swanson et al., 2001, Clark & Swanson, 2005, Torgerson et al., 2002). While it may be predicted that rapid evolution of SFPs might be more pronounced in systems where females have multiple mates (such as *G. bimaculatus*) than those that are monogamous, this expected pattern was not observed for a study

of 18 candidate SFPs in butterflies, where monogamy was unexpectedly linked to fast evolution of SFPs, perhaps due to relaxed selective constraints (Walters & Harrison, 2011). Research on SFPs in more diverse insects with well-described mating biology are thus needed (Walters & Harrison, 2011). *G. bimaculatus* has high female polyandry, complete sperm mixing, and exhibits extensive pre- and post-mating female choice (Simmons, 1986, Morrow & Gage, 2001). Using the recently available list of 134 SFPs in *D. melanogaster* (shown in Table S7, (Sepil et al., 2019)), we found that only 20 genes had identifiable putative orthologs in *G. bimaculatus* genes (14.9%). This is much lower than the 64.5% genome-wide rate of putative ortholog detection between these two species (Chi-square with Yates's correction $P<0.001$). Thus, the lack of putative SFP orthologs is consistent with especially rapid evolution (Tautz & Domazet-Loso, 2011, Haerty et al., 2007) of the SFP genes following the divergence of the lineages leading to *D. melanogaster* and *G. bimaculatus*.

Among the 20 putative *G. bimaculatus* SFP genes, seven were included among the subset of 7,220 genes with between-species orthologs in *Gryllus* (Table 4; note that none of these were among the 30 accessory gland-specific genes reported above). It has been inferred that SFPs tend be produced in insect accessory glands, as well as in the testis or male somatic reproductive system tissues (Sepil et al., 2019). Indeed, we found that each of these seven putative *Gryllus* SFPs exhibited expression within the testis, male somatic reproductive system, and the male accessory glands (between 0.2 to 1392.5 FPKM depending on tissue, with one exception, testis for GBI_14450-RA FPKM=0, Table 4). Significantly, for each of these seven putative cricket SFPs, we also found that the dN/dS values were consistently well above the median observed for all studied genes in the genome (which was 0.115 across all 7,220 genes, shown in Fig. 3A). Specifically, the values were 0.149 (*Phm*), 0.220 (*Npc2b*), 0.230 (*Ggt-1*), 0.250 (*regucalcin*), 0.287 (*Spn28F)*, 0.344 (*Spn38F*) and 2.48 (*Obp56g*) (Table 4). Thus, the putative SFPs in the crickets studied here have evolved very rapidly, a feature shared with the SFPs that have been studied in the fellow insect *D. melanogaster* (Sepil et al., 2019, Haerty et al., 2007). It should be noted that while we consider it unlikely, we cannot exclude the possibility that some accessory gland or SFP CDS may be expressed at extremely low levels in the *G. assimilis* tissue types used for RNA-seq, causing an apparent absence of orthologs to *G. bimaculatus* in that assembly. However, we consider this unlikely given the number of tissues we assessed, including the male accessory glands (Table S2). Moreover, this would not explain the apparent paucity of *G. bimaculatus* SFP orthologs relative to those in the *D. melanogaster* genome. Thus, we suggest the absence is best explained by rapid divergence that obscures ortholog detection, and/or from gene losses or gains (Tautz & Domazet-Loso, 2011, Haerty et al., 2007).

A role of positive selection for at least one SFP gene in *Gryllus* is supported by the fact that the dN/dS value was >1 (was 2.5, Table 4) for the odorant binding SFP protein *Obp56g*. In *D. melanogaster*, *Obp56g* was first recognized as an SFP using proteomics of seminal fluid in mated females (Findlay et al., 2008), was later affirmed as a protein stored in male reproductive tissues (Takemori & Yamamoto, 2009) (which we have confirmed also express this gene in crickets: (Table 4)), and was stringently verified as an SFP by Sepil and colleagues (2019).

Branch-site analysis for *G. bimaculatus*

Three-way reciprocal BLASTX of *Laupala kohalensis* to *G. bimaculatus* and *G. assimillus* yielded 4,523 genes with putative orthologs. Using free-ratio branch analyses of the three species, we found dN was largely unsaturated for the *L. kohalensis* branch, with a median of 0.10. However, dS values were particularly high (median=3.3), suggesting a high mutation rate in this organism. Including only genes with dN and dS <3 yielded 1,933 genes with confidence orthologs in *L. kohalensis* (26.7% of the 7,220 genes with *G. bimaculatus* and *G. assimillus* orthologs). This conservative approach favors study of the slowly evolving genes in each sex-biased category. We found instances of positive selection at specific sites in the *G. bimaculatus* branch for sex-biased genes from all studied tissue types (2XlnL P <0.05, Table S8). For instance, we found 11.8%, 9.6% and 10.9% of studied genes exhibited positive selection for ovary-biased$_{TSSB}$, testis-biased$_{TSSB}$ and universally unbiased genes (2XlnL P per gene <0.05; Table S8). The use of conserved genes, however, biases these testis-biased estimates of positive section downward (as fast evolving genes are excluded more often: 23.7% of testis-biased genes had three-way orthologs, versus 29.8% for ovary-biased genes). Further, while the number of genes, and thus three-way orthologs, were uncommon outside the gonads (N=4-33 depending on tissue; Table S8), we found that more than three times as many female-biased than male-biased somatic reproductive system genes exhibited branch-site selection (18.2% versus 5%; but this was not statistically significant, Chi-square P=0.17, Table S8), suggesting that this narrowed level of analysis (branch-site analysis of conserved genes), may concur with the notion that some genes from the female reproductive tract and/or spermathecae, which store sperm after mating, tend to evolve adaptively due to sexual selection pressures (Swanson et al., 2004, Prokupek et al., 2008). Future studies using more closely related cricket genomes as data emerge will be needed to enhance the power of detecting branch-site positive selection using branch-site analysis.

References

Andres, J. A., Larson, E. L., Bogdanowicz, S. M. & Harrison, R. G. 2013. Patterns of transcriptome divergence in the male accessory gland of two closely related species of field crickets. *Genetics* **193**: 501-13.

Andres, J. A., Maroja, L. S., Bogdanowicz, S. M., Swanson, W. J. & Harrison, R. G. 2006. Molecular evolution of seminal proteins in field crickets. *Mol Biol Evol* **23**: 1574-84.

Assis, R., Zhou, Q. & Bachtrog, D. 2012. Sex-biased transcriptome evolution in Drosophila. *Genome Biol Evol* **4**: 1189-200.

Clark, N. L. & Swanson, W. J. 2005. Pervasive adaptive evolution in primate seminal proteins. *PLoS Genet* **1**: e35.

Findlay, G. D., Yi, X., Maccoss, M. J. & Swanson, W. J. 2008. Proteomics reveals novel *Drosophila* seminal fluid proteins transferred at mating. *PLoS Biol* **6**: e178.

Haerty, W., Jagadeeshan, S., Kulathinal, R. J., Wong, A., Ram, K. R., Sirot, L. K., Levesque, L., Artieri, C. G., Wolfner, M. F., Civetta, A. & Singh, R. S. 2007. Evolution in the fast lane: rapidly evolving sex-related genes in Drosophila. *Genetics* **177**: 1321-1335.

Harrison, P. W., Wright, A. E., Zimmer, F., Dean, R., Montgomery, S. H., Pointer, M. A. & Mank, J. E. 2015. Sexual selection drives evolution and rapid turnover of male gene expression. *Proc Natl Acad Sci U S A* **112**: 4393-8.

Huang da, W., Sherman, B. T. & Lempicki, R. A. 2009. Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. *Nature Protocols* **4**: 44-57.

Morrow, E. H. & Gage, M. G., H 2001. Sperm competition experiments between lines of crickets producing different sperm lengths. *Proceedings of the Royal Society of London. Series B: Biological Sciences* **268**: 2281-6.

Prokupek, A., Hoffmann, F., Eyun, S. I., Moriyama, E., Zhou, M. & Harshman, L. 2008. An evolutionary expressed sequence tag analysis of *Drosophila* spermatheca genes. *Evolution* **62**: 2936-47.

Sepil, I., Hopkins, B. R., Dean, R., Thezenas, M. L., Charles, P. D., Konietzny, R., Fischer, R., Kessler, B. M. & Wigby, S. 2019. Quantitative Proteomics Identification of Seminal Fluid Proteins in Male *Drosophila melanogaster*. *Mol Cell Proteomics* **18**: S46-S58.

Seppey, M., Manni, M. & Zdobnov, E. M. 2019. BUSCO: Assessing Genome Assembly and Annotation Completeness. *Methods Mol Biol* **1962**: 227-245.

Simmons, L. W. 1986. Female choice in the field cricket *Gryllus bimaculatus (*De Geer). *Animal Behavior* **34**: 1463-1470.

Swanson, W. J., Clark, A. G., Waldrip-Dail, H. M., Wolfner, M. F. & Aquadro, C. F. 2001. Evolutionary EST analysis identifies rapidly evolving male reproductive proteins in Drosophila. *Proceedings of the National Academy of Sciences of the United States of America* **98**: 7375-7379.

Swanson, W. J., Wong, A., Wolfner, M. F. & Aquadro, C. F. 2004. Evolutionary expressed sequence tag analysis of *Drosophila* female reproductive tracts identifies genes subjected to positive selection. *Genetics* **168**: 1457-65.

Takemori, N. & Yamamoto, M. T. 2009. Proteome mapping of the Drosophila melanogaster male reproductive system. *Proteomics* **9**: 2484-93.

Tautz, D. & Domazet-Loso, T. 2011. The evolutionary origin of orphan genes. *Nature Reviews Genetics* **12**: 692-702.

Torgerson, D. G., Kulathinal, R. J. & Singh, R. S. 2002. Mammalian sperm proteins are rapidly evolving: evidence for positive selection in functionally diverse genes. *Molecular Biology and Evolution* **2002**.

Walters, J. R. & Harrison, R. G. 2011. Decoupling of rapid and adaptive evolution among seminal fluid proteins in heliconius butterflies with divergent mating systems. *Evolution* **65**: 2855-71.

Whittle, C. A. & Extavour, C. G. 2019a. Contrasting patterns of molecular evolution in metazoan germ line genes. *BMC Evol Biol* **19**: 53 (doi: 10.1186/s12862-019-1363-x).

Whittle, C. A. & Extavour, C. G. 2019b. Selection shapes turnover and magnitude of sex-biased expression in *Drosophila* gonads. *BMC Evol Biol* **19**: 60 (doi: 10.1186/s12862-019-1377-4).

Ylla, G., Nakamura, T., Itoh, T., Kajitani, R., Toyoda, A., Tomonari, S., Bando, T., Ishimaru, Y., Watanabe, T., Fuketa, M., Matsuoka, Y., Barnett, A. A., Noji, S., Mito, T. & Extavour, C. G. 2021. Insights into the genomic evolution of insects from cricket genomes. *Communications Biology* **in press**.

Zhang, Y., Sturgill, D., Parisi, M., Kumar, S. & Oliver, B. 2007. Constraint and turnover in sex-biased gene expression in the genus Drosophila. *Nature* **450**: 233-7.